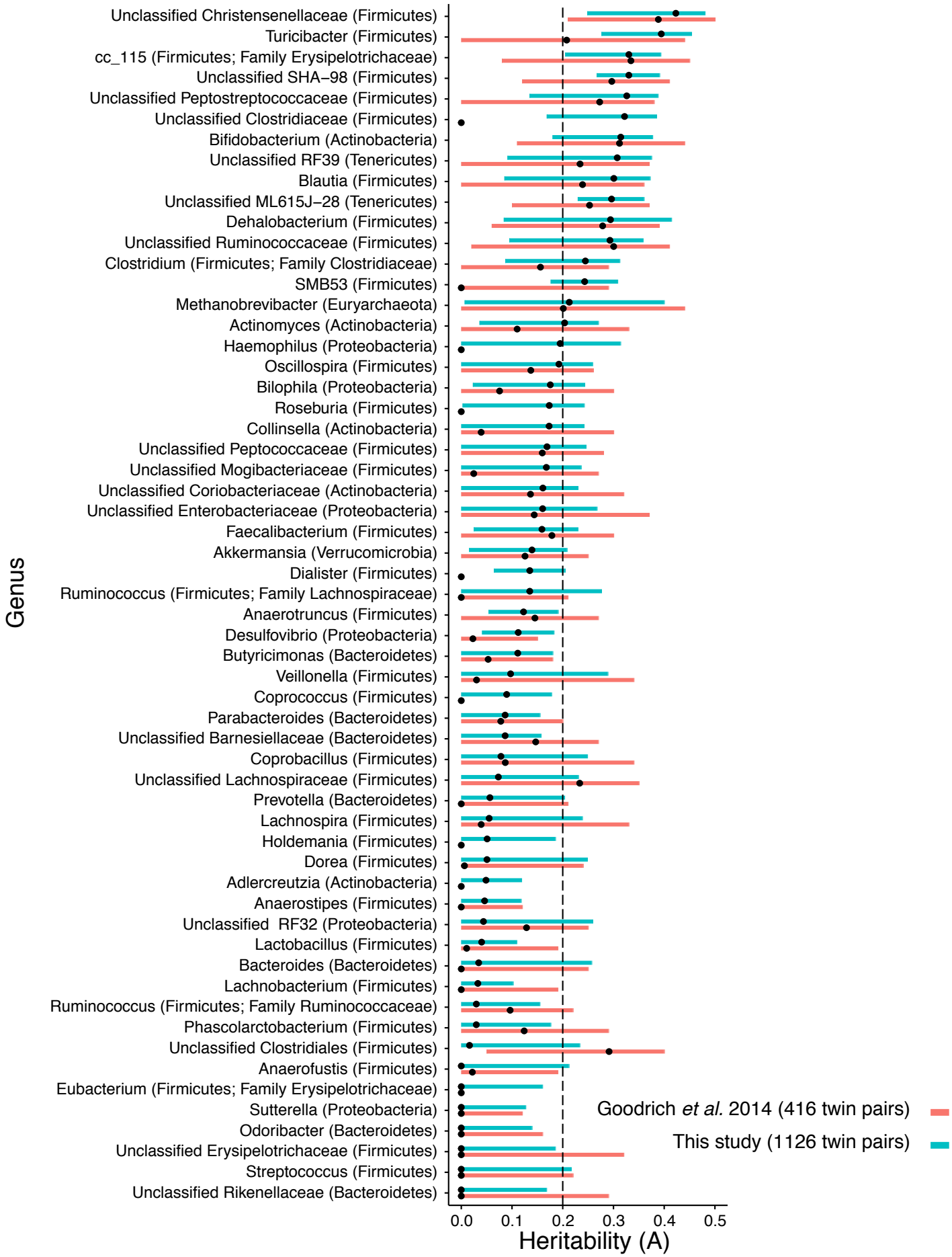

Figure S1. Heritability estimates and confidence intervals for all common genera. Related to Figure 1 and Table S1. Each point represents the heritability estimated as the proportion of variance in the microbial abundances (OTU abundances collapsed by taxonomic classification) that can be attributed to genetic effects (A). The bars show the 95% confidence intervals around the heritability estimates. Bars are colored by the dataset: Pink indicates heritability estimates reported in Goodrich et al. 2014 (171 MZ, 245 DZ) and Blue indicates heritability estimates using an additional 710 twin pairs (Blue; 637 MZ, 489 DZ twin pairs). The figure includes only genera that are present in at least 50% of the TwinsUK participants in the increased dataset.

Table S1. Heritability estimates for all 945 taxa. Related to Figures 1 and S1. This table contains the A, C and E estimates and their 95% confidence intervals from the ACE model (A = additive genetic influence; C = shared environmental influence; E = nonshared environmental influence) and *P*-value for A (likelihood ratio test comparing the ACE model to the CE model).

Table S2. Genetic and environmental contribution to Alpha- and Beta- diversity metrics. Related to Figure 2.

	Trait	A (95% CI)	*P*-value	C (95% CI)	E (95% CI)	ICC_{MZ}	ICC_{DZ}	Falconer h^2
Alpha Diversity	PD whole tree	0.37 (0.17-0.44)	0.0004	0.01 (0-0.19)	0.62 (0.56-0.68)	0.39	0.19	0.40
	Shannon	0.32 (0.26-0.38)	0.0005	0 (0-0.13)	0.68 (0.62-0.74)	0.33	0.14	0.38
	Observed species	0.3 (0.24-0.37)	0.0018	0 (0-0.14)	0.7 (0.63-0.76)	0.31	0.13	0.36
Beta Diversity	Unweighted UniFrac PC1	0.47 (0.29-0.58)	2.1×10^{-07}	0.06 (0-0.21)	0.47 (0.42-0.53)	0.53	0.29	0.48
	Unweighted UniFrac PC2	0.25 (0.09-0.31)	0.0053	0 (0-0.1)	0.75 (0.69-0.82)	0.25	0.10	0.32
	Unweighted UniFrac PC3	0.13 (0-0.32)	0.2559	0.13 (0-0.28)	0.74 (0.67-0.81)	0.26	0.19	0.15
	Weighted UniFrac PC1	0 (0-0.16)	1.0000	0.2 (0.06-0.26)	0.8 (0.74-0.85)	0.20	0.21	-0.03
	Weighted UniFrac PC2	0.15 (0-0.22)	0.1094	0 (0-0.15)	0.85 (0.78-0.92)	0.17	0.05	0.23
	Weighted UniFrac PC3	0.11 (0-0.29)	0.3388	0.12 (0-0.25)	0.77 (0.7-0.84)	0.23	0.17	0.12
	Bray Curtis PC1	0.15 (0-0.36)	0.1597	0.17 (0-0.33)	0.68 (0.61-0.75)	0.32	0.25	0.15
	Bray Curtis PC2	0.08 (0-0.3)	0.4830	0.17 (0-0.28)	0.75 (0.69-0.82)	0.26	0.20	0.12
	Bray Curtis PC3	0.17 (0-0.24)	0.0708	0 (0-0.16)	0.83 (0.76-0.9)	0.18	0.07	0.24

For each Metric, we report the A, C and E estimates from the ACE model (A = additive genetic influence; C = shared environmental influence; E = nonshared environmental influence) and *P*-value for A (likelihood ratio test comparing the ACE model to the CE model). Intraclass coefficient of correlation for MZ (ICC_{MZ}) and DZ (ICC_{DZ}), Falconer ($h^2 = 2*(ICC_{MZ}-ICC_{DZ})$) heritability.

Table S3. List of all genes tested in candidate gene sets. Related to Tables 1, 2 and Table S4. This table contains all genes tested in the candidate gene sets and which gene set they belong to

Table S4. List of all SNPs tested in candidate gene sets. Related to Tables 1, 2 and S3. This table contains all SNPs tested in the candidate sets and which candidate set they belong to

Table S5. GEMMA GWAS hits significant at genomewide FDR of 0.2. Related to Figure 4. This table contains all taxa ~ SNP associations that had a genomewide FDR correction <0.2. Only most associated SNP in each locus is reported.

Table S6. PrediXcan hits significant at a threshold of $P < 5 \times 10\text{-}8$. Related to Figure 4. This table contains all taxa ~ imputed gene expression associations that were more significant that $P < 5 \times 10\text{-}8$.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Sample collection and 16S rRNA gene data analysis

The entire sample set consists of 3,261 fecal samples from 2,731 individuals (530 individuals provided two or more samples at different time points) including 489 DZ twin pairs, 637 MZ twin pairs. The average age was 59 years (range 18–89 years) and 89% were female. This sample set includes the 1,081 samples collected for Goodrich 2014 (European Nucleotide Archive, European Bioinformatics Institute, with accession numbers ERP006339 and ERP006342), samples previously described by Jackson et al. (2015) and an additional set of newly collected and processed samples. For the newly collected samples, DNA extraction, amplification of the V4 hypervariable region of the 16S rRNA gene (primers 515F and 806R), purification and pooling was performed on all fecal samples as previously described (Goodrich et al., 2014). The pooled amplicons were sequenced using the Illumina Miseq platform with 2x250bp paired-end sequencing.

Mate-pair merging, de-multiplexing, quality control and OTU picking were performed using QIIME version 1.8 (Quantitative Insights Into Microbial Ecology; Caporaso et al., 2010). Matching paired-end sequences (mate-pairs) were merged using fastq-join through QIIME with a minimum overlap of 200 base pairs (Aronesty, 2011). Quality filters were used to remove sequences containing uncorrectable barcodes, ambiguous bases, or low quality reads (Phred quality scores ≤ 25) resulting in a total of 302,554,236 sequences (an average of 90,100 sequences per sample).

Sequences were clustered into OTUs using the default pick open reference OTU pipeline in QIIME and the Greengenes August 2013 97% identity OTU representative sequences and taxonomy as the reference database (McDonald et al., 2012). β-diversity (unweighted and weighted UniFrac (Lozupone et al., 2007) metrics and Bray-Curtis Dissimilarity (Bray and Curtis, 1957)) was calculated with an OTU table rarefied to 10,000 sequences per sample and Principal Coordinate Analysis (PCoA) was performed on these distance matrices. α-diversity (Faith's phylogenetic diversity (Faith, 1992), observed species, and Shannon diversity) was calculated from 100 iterations using a rarefaction of 10,000 sequences per sample.

Taxa filtering and covariate regression

Summaries of the taxonomic distributions were generated by collapsing OTU counts based upon their taxonomic classification at five levels from genus to phylum. Due to the difference in the taxonomy classification from Goodrich et al. (Goodrich et al., 2014) the collapsed taxonomy classification of k__Bacteria; p__Firmicutes; c__Clostridia; Other; Other; Other in this dataset is actually the Greengenes OTU 836693 which has the taxonomy k__Bacteria; p__Firmicutes; c__Clostridia; o__SHA-98; f__; g__; s__. To maintain consistency with the original dataset we renamed it to SHA-98. For downstream analyses the abundances were first filtered by percentage of individuals sharing the taxa. Any taxon (OTU or summarized by taxonomy) shared by less than 50% of the individuals in the study (less than 50% of the counts are non-zero) was excluded from further analyses. Relative abundances were Box-Cox transformed using the equation

$$y^{(\lambda)} = (y^{\lambda} - 1) / \lambda$$

λ was optimized using the PowerTransform command implemented in the R package 'car' and before normalizing to sequencing depth an offset of one was added to handle zero counts. A multiple linear regression was performed where the transformed taxa abundances were regressed on covariates (the number of 16S rRNA gene sequences per sample, age, gender, shipment date, collection method (postal or in person), and ID of technician performing DNA extraction). The residuals from this regression were then used for the heritability estimates, correlations with alpha- and beta-diversity, and the genome-wide association studies. After filtering we obtained 945 microbiome traits, including 782 OTUs, 73 genera, 43 families, 21 orders, 17 classes, and 9 phyla.

Alpha-diversity was also normalized using the Box-Cox transformation. Age, gender, shipment date, collection method (postal or in person), and ID of technician performing DNA extraction were regressed from the normalized alpha-diversity and the first three PCs from the PCoA of the beta-diversity matrices and the residuals were used in subsequent analyses.

Correlations between diversity metrics and taxa

Given the hierarchical nature of taxonomy, some of the higher taxonomic levels consist of only a single OTU or genus. To prevent redundancy, if any trait had a correlation coefficient of 0.9 or greater with another trait, only the trait with the most informative taxonomic classification was kept in the analysis. Pairwise pearson correlations were calculated between taxa abundances and alpha- and beta-diversity residuals described above. A heatmap of the correlations, made using the package pheatmap in R, shows only taxa with $| r | > 0.5$ to any of the three alpha-diversity metrics or the 9 beta-diversity PCs.

Heritability calculations

The ACE model (Eaves et al., 1978) estimates the additive genetic (A), common environment (C) and unique environment (E) components of the trait variance. The heritability is defined as the proportion of total variance that is due to genetic effects (A/V). The estimates were obtained using the structural equation modeling (SEM) software OpenMx (Boker et al., 2011). OpenMx was also used to calculate the CE submodel. The significance of A was determined by using the likelihood ratio test to comparing the fit of the CE model to the fit of the ACE model. In the case of longitudinal samples from the same individual, the first sample collected was used. To provide multiple testing correction for the heritability analysis, we used the Benjamini-Hochberg algorithm in R to correct for all 945 traits tested.

We also estimated the heritability using the MZ and DZ intraclass correlation coefficients (ICCs) to confirm the SEM method was behaving as expected. Heritability was estimated as $2*(ICC_{MZ}- ICC_{DZ})$ (i.e. Falconer's formula). All ICC calculations were generated with the 'icc' command from the R package 'irr'.

Host genetic association analyses

For MZ twin pairs with only one twin genotyped, that genotype information was used for both twins. SNPs were removed if they had a minor allele frequency below 5%, a genotyping rate below 95% or significant deviation from Hardy-Weinberg Equilibrium (HWE; $p<0.001$). The final number of SNPs used in the genome-wide association analyses was 1,300,091. GEMMA (v0.94) was used to calculate a kinship matrix between individuals and to perform a GWAS on the residuals (described above) for each taxon, alpha-diversity metric, and the first 3 PCs from PCoA on each of the beta-diversity metrics. The *P*-values from the likelihood ratio test in GEMMA were adjusted using the Benjamini-Hochberg correction on the *P*-values for the 1,300,091 SNPs (written as BH adjusted *P*-value in the results section) to control for false-positive error rates deriving from multiple testing at the genome-wide level.

Candidate gene and SNP analysis

The list SNPs/Genes for the candidate sets were obtained from the associations found by the references listed in **Table 1**. If the reported SNPs were not present in the imputed genotype data, a proxy SNP was chosen using the SNP Annotation and Proxy Search (SNAP) tool with the CEU population panel from the 1000 Genomes Pilot 1 data (Johnson et al., 2008). The proxy SNP with the highest r^2 and closest proximity to the reported SNP was then used. For the gene set analysis, all SNPs within 5kb of the reported gene were examined. To reduce the burden of multiple testing, we examined the collapsed taxa (not the OTUs) with heritability of $A>0.2$. Given the redundancy between the levels of taxonomy, if a taxon had a correlation coefficient of 0.9 or greater with another heritable taxon, only the one with the most informative taxonomic classification was kept in the analysis; 20 taxa passed this filter.

To calculate a *P*-value for each gene set, we compared the minimum *P*-value for all SNP x taxa associations (calculated by GEMMA as described above) to the distribution of minimum *P*-values for 1000 permutations of the taxa residuals. The taxa residuals were permuted across individuals and the the correlation structure of the phenotype data was preserved. Specifically, the permutations were restricted to preserve the twin structure and zygosity: MZ-twin pairs were permuted within the twin pair and assigned to another MZ-pair and the same procedure was applied for DZ-pairs.

We also report the minimum BH adjusted *P*-value within a gene set (correction for all SNPs in the gene set and the 20 heritable traits tested). The permutation *P*-value correlates with the min BH adjusted *P*-value within a gene set.

***Bifidobacterium* validation in the Hutterite dataset**

All 16S rRNA gene sequencing analysis and genotyping is described in the original study (Davenport et al., 2015). Briefly, the dataset consists of fecal 16S rRNA gene sequence data from the Hutterites sampled during two seasons (summer n = 91; winter n = 93; seasons combined n = 127). The relative abundance of *Bifidobacterium* was fit to a standard normal distribution across individuals, and age, sex and date of collection were regressed from the normalized *Bifidobacterium* relative abundance. We used PRIMAL (Livne et al., 2015), an in-house pedigree-based imputation algorithm, to first phase and then impute all variants in the genome sequences of 98 Hutterites to 1,317 Hutterites who were previously genotyped on Affymetrix arrays (Yao et al., 2014). This imputation method assigns genotypes with >99% accuracy (Livne et al., 2015). Genotype at SNP rs4988235 was imputed to 91.3% of individuals in the sample, including 121/127 for the seasons combined microbiome analysis, 89/91 in summer microbiome analysis, and 88/93 in the winter microbiome analysis. GEMMA (v0.94) was used to perform the

association between the *Bifidobacterium* residuals with the SNP rs4988235, including the relatedness matrices as described by Davenport et al. (2015).

GWAS for beta-diversity measures

The tool microbiomeGWAS (Hua et al., 2015) was downloaded from https://github.com/lsncibb/microbiomeGWAS on 2/24/16. Only one twin per family included in this analysis, for a total sample size of 1,248 individuals. microbiomeGWAS was run for three different beta-diversity metrics (Bray Curtis dissimilarity, weighted UniFrac distance, and unweighted UniFrac distance), including the following covariates in the model: the number of 16S rRNA gene sequences per sample, age, gender, shipment date, collection method (postal or in person), and ID of technician performing DNA extraction.

Imputed gene expression analysis

PrediXcan software was downloaded from https://github.com/hakyimlab/PrediXcan/tree/master/Software and database files from https://app.box.com/s/gujt4m6njqjfqqc9tu0oqgtjvtz9860w on 2/12/2016 (Gamazon et al., 2015). Only database files designated as "ElasticNet.0.5.db" were considered in the analysis (40 tissues total). Genotype files in binary plink format were converted to dosage format using convert_plink_to_dosage.py and gene expression was inferred for each tissue using predict_gene_expression.py. GEMMAv0.94 was used to perform association analysis as described above. Thresholds of significance were set considering study-wide significance (0.05/(40 tissues x 945 taxa x 338,372 genes)), tissue-wide significance (0.05/(945 x # genes imputed in that tissue)), or at 5 $x10^{-8}$.

Cross-study comparisons of taxa influenced by host genetics

The reported heritability estimates were obtained from the following tables in their respective studies: Davenport et al. 2015 - Proportion of variance explained estimates ("chip heritability") and standard errors in Table S2; Org et al. 2015 - Heritability estimates in Table S3; O'Connor et al. 2014 - Heritability estimates in Table 1

The reported taxa with QTL/GWAS results were obtained from the following tables in their respective studies: Davenport et al. 2015 - Taxa associated with at least one SNP in Table S7; Blekhman et al. 2015 - Stool GWAS results in Table S5; Benson et al. 2010 - QTL results in Table S5; Leamy et al. 2014 - QTL results in Table 1; McKnite et al. 2012 - QTL results in Table 2; Org et al. 2015 - GWAS results in Table S6

To determine if the taxon is not observed in the study or was excluded from the original analysis the following tables or figures were used: Davenport et al. 2015 - Table S2; Blekhman et al. 2015 - Figure S18; Benson et al. 2010 - Table S2; McKnite et al. 2012 - Figure 1; Leamy et al. 2014 - Additional file 1 - Only Core measurable microbiota (CMM) that have a name exactly matching the taxa listed in this table are considered; O'Connor et al. 2014 - Table S1; Org et al. 2015 - Table S3.

SUPPLEMENTAL REFERENCES

Aronesty, E. (2011). ea-utils : "Command-line tools for processing biological sequencing data" (http://code.google.com/p/ea-utils).

Bray, J.R., and Curtis, J.T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecol Monogr *27*, 326-349.

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T.*, et al.* (2011). OpenMx: An Open Source Extended Structural Equation Modeling Framework. Psychometrika *76*, 306-317.

Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. Biol Conserv *61*, 1-10.

Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., and de Bakker, P.I.W. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics *24*, 2938-2939.

Livne, O.E., Han, L., Alkorta-Aranburu, G., Wentworth-Sheilds, W., Abney, M., Ober, C., and Nicolae, D.L. (2015). PRIMAL: Fast and accurate pedigree-based imputation from sequence data in a founder population. PLoS

Comput. Biol. *11*, e1004139.

Lozupone, C.A., Hamady, M., Kelley, S.T., and Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol *73*, 1576-1585.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J *6*, 610-618.

Yao, T.-C., Du, G., Han, L., Sun, Y., Hu, D., Yang, J.J., Mathias, R., Roth, L.A., Rafaels, N., Thompson, E.E., et al. (2014). Genome-wide association study of lung function phenotypes in a founder population. J. Allergy Clin. Immunol. *133*, 248–255.e1–e10.